Latent disconnectome prediction of long-term cognitive symptoms in stroke

Lia Talozzi (✉ liatalozzi@gmail.com)

  CNRS, University of Bordeaux   https://orcid.org/0000-0003-1827-0407

Stephanie Forkel

  CNRS, Unviersity of Bordeaux

Valentina Pacella

  CNRS, Unviersity of Bordeaux

Victor Nozais

  CNRS, Unviersity of Bordeaux

Maurizio Corbetta

  University of Padua

Parashkev Nachev

  University College London

Michel Thiebaut de Schotten

  CNRS, University of Bordeaux

Article

Keywords:

Abstract

Stroke significantly impacts quality of life. However, the long-term cognitive evolution in stroke is poorly predictable at the individual level. There is an urgent need for a better prediction of long-term symptoms based on acute clinical neuroimaging data. Previous works have demonstrated a strong relationship between the location of white matter disconnections and clinical symptoms. However, rendering the entire space of possible disconnections-deficit associations optimally surveyable will allow for a systematic association between brain disconnections and cognitive-behavioural measures at the individual level. Here we present the most comprehensive framework, a composite morphospace to predict neuropsychological scores one year after stroke. Linking the latent disconnectome morphospace to neuropsychological outcomes yields biological insights available as the first comprehensive atlas of disconnectome-deficit relations across 86 neuropsychological scores. Out-of-sample prediction derived from this atlas achieved average accuracy over 80%, which is higher than any other framework. Our novel predictive framework is available as an interactive web application, the *disconnectome symptoms discoverer* (http://disconnectomestudio.bcblab.com), to provide the foundations for a new and practical approach to modelling cognition in stroke. Our atlas and web application will reduce the burden of cognitive deficits on patients, their families, and wider society while also helping to tailor personalized treatment programs and discover new targets for treatments. We expect the range of assessments and the predictive power of our framework to increase even further through future crowdsourcing.

Introduction

The fidelity of lesion-deficit models depends not only on the quality of the data but also on the underlying theoretical framework. Together they produced evidence of a relationship between the location of brain lesions and clinical symptoms such as visuospatial neglect [1-3], aphasias [4-6], apraxias [7,8] or motor anosognosia [9,10] amongst others. Recently, the associations between anatomical white matter networks and clinical presentations revealed that there is no one-to-one relationship between structures and clinical presentation, as different lesions can cause the same functional impairments [11,12]. One example would be that a stroke in the middle or posterior cerebral artery may lead to visuospatial neglect [13], just like different perisylvian white matter disconnections can lead to aphasia [11]. Hence, the current methodologies do not capture the potential overlap between brain signatures and clinical manifestations nor the distributed nature of their neural substrate, now familiar from network analyses of functional imaging data [14]. Therefore, a comprehensive framework that would systematically associate brain disconnections with cognitive-behavioural assessments is needed for accurate precision medicine [15-19].

Modelling distributed relations is computationally expensive and requires large scale data. With advances in data modelling and the availability of databases, tackling the high complexity of clinical-anatomical relationships is now conceivable. Beneath the surface complexity there may lie a simpler order that can be described within a compacted representational space. As such, dimensionality reduction algorithms allow defining low-dimensional spaces that can embed multivariate data. In embedding spaces, also

known as morphospaces [20,21], patients with similar features cluster together while diverging features are placed apart [12,22]. Morphospaces render lesion-deficit relations more easily surveyable. Hence, specific brain features can define territories in a morphospace and help predict symptoms and brain pathologies, similar to typical machine learning approaches [23,24]. Artificial intelligence (AI) has recently progressed in modelling the association of symptom severity with medical imaging modalities, e.g., reaching high accuracy and sensitivity in the characterisation of tumour tissues [25]. However, AI models need to be refined with a broader spectrum of clinically practical endpoints, including neuropsychological measures. The next challenge will be making AI patient-centric for a more effective deployment into the clinical routine and to efficiently benefit patients' quality of life [26].

To drive the realisation of this challenge forward, we propose a modelling approach that employs a morphospace to predict neuropsychological assessments of one of the most common neurological disorders: stroke [27]. We first mapped the distribution of 1333 brain disconnection patterns in stroke — the disconnectome morphospace. A second dataset (training set) with rich neuropsychological measures 1-year after stroke was imported into this disconnectome morphospace. This second dataset enriched the morphospace with clinical symptoms obtained from 86 neuropsychological assessments. An out-of-sample "validation set" with the same neuropsychological data served to assess prediction accuracy. This procedure, hereafter referred to as *disconnectome symptoms discoverer* (DSD), reliably predicted the performance of patients with an average accuracy > 80%. To make the DSD tool readily available to the clinical-academic community and facilitate its incorporation into the clinic, we provide an open-access web application (http://disconnectomestudio.bcblab.com), in which individual disconnection patterns can be uploaded to predict the expected 1-year neuropsychological scores. The web application will be interactively updated, thanks to future crowdsourcing, informing the DSD model with any newly available datasets.

Results

The disconnectome morphospace

The first dataset (N=1333 stroke lesions [28]; see Supplementary Table 1) was processed to obtain disconnectome maps. Disconnectome maps quantify the pattern of connections interrupted by each lesion based on the high-resolution tractography of a healthy population [12,29,30]. Subsequently, the Uniform Manifold Approximation and Projection (UMAP) [31] method was used to embed disconnection complexity. A latent 2 dimensions configuration of the disconnectome maps was obtained. Figure 1 indicates that patients' disconnectome profiles distribute based on lesion location and commonly disconnected tracts. For instance, patients with major left or right hemisphere disconnections were embedded in the right and left half of the morphospace, respectively. Similarly, patients with posterior or anterior disconnections were localised at the top or the bottom of the embedded space. Patients with a prominent disconnection of the inferior-fronto occipital fasciculus (IFOF) located at the bottom left and right extremities of the morphospace while corticospinal (CST) and arcuate (AF) disconnections were

relatively more central. Hence the morphospace appropriately segregated the different profiles of disconnection of the classic tract [32,33].

The composite morphospace

The extent to which the disconnectome morphospace can predict different neuropsychological performances is currently unknown. To answer this question, we took advantage of the second independent dataset of stroke patients (N=119 stroke lesions [14]; see Supplementary Table 1) that was extensively explored with standard neuropsychological assessments (N=86, see Supplementary Table 2). For each patient of the second dataset, disconnectome maps were calculated and imported into the disconnectome morphospace using the UMAP defined transformation. To tackle uncertainty, patient coordinates in the morphospace were spatially smoothed (see methods). In so doing, each patient's coordinates in the disconnectome morphospace were converted into probabilities of localisation. A Pearson correlation approach was then used to estimate the association between each morphospace coordinate and a neuropsychological performance (see Supplementary Figure 1 for more details). Figure 2 indicates that a medium to large effect size association (all |r| >0.2) existed between territories in the disconnectome morphospace and neuropsychological scores (Figure 2a-c). Importantly, for some scores, multiple clusters in the disconnectome morphospace, corresponding to different disconnection profiles, apparently led to the same neuropsychological impairment. This confirmed that no one-to-one relationship exists between lesion of structures and clinical disorders, and likewise, different locations of brain damage can lead to the same functional impairment. To avoid simple linear association between the morphospace coordinates magnitude and neuropsychological scores, patients' probabilities of localisation in clusters of significance were modelled by a principal component analysis (later referred to as spatial PCA). For each patient, the first three-component of the spatial PCA were entered into a multiple regression analysis to predict single-patient neuropsychological scores 1 year after symptom onset. The multiple regressions created equations, modelling the relationship between each patient's potential localisation in the disconnectome morphospace (i.e., as defined by the three first components of the spatial PCA) and their neuropsychological scores. In so doing, we obtained a composite morphospace that takes advantage of the joint strengths of the two datasets. The composite morphospace accurately and reliably predicted 83 out of 86 neuropsychological scores with a small to large effect size (see Supplementary Table 3).

Disconnectome morphospace component mapping

In the next level, we brought the score prediction results back to the neuroimaging space to explore the neuroanatomical patterns leading to symptoms. The first dataset was split in half (2 X 666 disconnectomes maps) to assess reproducibility. Latent patterns of predicted neuropsychological performances were statistically associated with brain disconnections maps of the two halves of the first dataset using voxelwise linear regressions. In doing so we obtained two sets of maps of brain disconnection for each neuropsychological score (see example in figure 2d-f and all maps together with their full discussion in Supplementary Material - Section C). We were able to produce a comprehensive

atlas of the brain disconnections associated with neuropsychological test scores and the statistical comparison of the two sets of maps indicated a good level of reproducibility (Pearson R = 0.82). Figure 3 summarises the highest statistical associations spanning from a medium ($0.25 > f^2 > 0.42$) to a high effect size ($0.42 < f^2$). The highest effect sizes were in the left hemisphere, particularly in the frontal lobe connections, indicating the strongest associations between these disconnections and neuropsychological scores (Figure 4a). Some areas can also be associated with multiple different neuropsychological scores. To summarise this information, we calculated a versability map that indicates how many neuropsychological scores can be predicted with a large effect size per volume unit of white matter (Figure 4b). The versatility maps revealed a clear asymmetry between the left and the right hemispheres. This lower effect size and higher versatility in the right hemisphere suggests that more work is required to finely measure and dissociate right hemisphere functions in neuropsychology.

Accuracy in predicting neuropsychological score at 1-year after stroke

To assess the accuracy of the predictions, data derived from a third independent dataset (20 stroke patients withheld from the original dataset[14]; see Supplementary Table 1) were projected into the morphospace. From there, equations derived from the composite morphospace were applied to predicted individual neuropsychological scores. Prediction accuracy was assessed as the difference between the observed and predicted scores, normalized by the maximum score (i.e., normalized prediction error; Figure 5). The profile of neuropsychological scores for single patients was predicted with an average accuracy of 84.3 ± 5.6 % while each test was predicted individually with an average accuracy of 83.9 ± 7 %. Overall, the prediction of two-thirds (N=65) of the tests was replicated in this third independent dataset with an accuracy >80% (Supplementary Table 4).

Disconnection Symptoms Discovery web application

To make this resource and method available for the clinical-research community, we deployed an interactive web application platform called *Disconnectome Symptoms Discovery - DSD* (http://disconnectomestudio.bcblab.com*). The DSD requires the input of brain lesions converted to disconnection maps and returns the expected 1-year neuropsychological scores for individual disconnectome maps (see the DSD user guide in the Supplementary Material Section E). The DSD tool prediction model relies on the databases presented in this study that can be updated on-demand with new neuropsychological assessments and patients' disconnectomes.

Discussion

Applying state-of-the-art data embedding methods we succeeded in combining complementary databases of stroke patients and produced an atlas of neuropsychological scores associated with brain disconnections. This atlas applied to an out-of-sample dataset accurately predicted 65 neuropsychological scores with an accuracy of over 80%. An openly available web application, the

disconnectome symptoms discoverer (DSD) capitalises on our methods and provides new anatomical insights into cognitive symptoms for researchers and clinicians.

Similar patterns of stroke-induced white matter disconnections were distributed close by in the embedding space comparably to other research fields using UMAP methods [31], e.g., single-cell genetic transcriptomes [34,35]. Therefore, the disconnectome morphospace acted as a reference to quickly import and summarize new stroke disconnections. Such embedded information allowed us to associate single-patient neuropsychological profiles at 1-year after a stroke with territories in the morphospace and profile of disconnection. By exploring white matter correlates systematically, we created a comprehensive atlas of the neuropsychological scores associated with brain disconnections. Classical functional associations were confirmed, e.g., the lateralisation of motor functions, the left perisylvian language network, the fronto-parietal attentional networks, or the right insula for sickness sensations. In addition, new insights on functioning and disconnection were reported, e.g., the callosum connectivity related to visual neglect, the cerebellum hub for visuospatial memory, and the lingual gyrus for verbal memory (for individual results and discussion see Supplementary Material - Section C).

The atlas allowed for the evaluation of acute MRI scans to predict long-term stroke symptom severity. These results indicate the suitability of the disconnectome model in predicting a wide range of functional performances and addressing a complete personalised, individual patient profile. This information will be a valuable resource in clinical settings, for example for the planning of personalized therapeutic and rehabilitation strategies. This is a step forward in comparison to many stroke AI methods that have a purely diagnostic purpose [36]. The DSD model has a prognostic vocation based on cross-modal data (neuroimaging input - neuropsychological outcome prediction).

However, predictions were not equally accurate across functions (see Supplementary Table 4). Three factors might explain these differences. First, some neuropsychological scores are more reliable than others in assessing performances [37]. Second, plasticity and interindividual variability might interact with recovery [11,38-40]. Third, the disconnectome model may not capture all the variance of brain injuries. Indeed, hypoperfusion [41] and hypometabolism [42] factors as well as acute imaging changes such as pseudonormalization [43] are not included.

Besides these limitations, the *disconnectome symptoms discoverer* (DSD) web application is a free and user-friendly web browser tool that only requires an internet connection. Instant software access and automatic updates make the word-wide-web the ideal media for clinical translations. The application of the DSD results can help personalized prognosis. Further, while our predictions were validated in an out-of-sample dataset, the DSD web application allows for a wider validation with crowdsourcing usage, through new dataset implementation. Hence, the DSD aims to benefit the researchers' understanding of brain functioning and the patient's treatments alike.

Methods

Bash and Python programming languages have been used for automatizing all the processing steps summarized in Supplementary Materials figure 1.

Stroke lesions. Lesion data were derived from two different centres. Dataset 1: (N=1333 strokes) MRI scans were acquired at University College London Hospitals [28]. Patient demographics were an average age of 6416 years [age range: 18-97 years] and 56% were men. This cohort was approved by the West London & GTAC Research Ethics Committee for consent use of fully anonymised data. Dataset 2 and dataset 3 were recruited at the School of Medicine of the Washington University in St. Louis and included both MRI and neuropsychological assessments [14]. All participants of datasets 2 and 3 provided informed consent following the Declaration of Helsinki (2013) and procedures established by the Washington University in Saint Louis Institutional Review Board. For dataset 2 (N=119 strokes), the patient average age was 5411 years [age range: 19-83 years] and 54% men. The average education level was 132.5 years, and 91% were right-handed participants. For dataset 3 (N=20 strokes), the average age was 5812 years [age range: 34-95 years], with 40% men; average education level was 142.6 years, and 85% were right-handed. Patient imaging parameters, inclusion criteria and demographics are reported in the Supplementary Materials Section A and lesion distributions across the three databases in the Supplementary Figure 2.

Neuropsychological scores. Neuropsychological scores were available for datasets 2 and 3. The details of each neuropsychological evaluation (grading, test battery, administration) are reported in the Supplementary Materials Section C. In brief, motor abilities (Section C.1) were assessed for upper limb hand grasping, gripping, pinching, grip strength, peg replacement, motion shoulder flexion, wrist extension, and lower limb walking. Language abilities (Section C.2) were assessed using picture naming, non-word repetition, commands, sentence reading, sentence comprehension, and semantic fluency. Visuospatial abilities (Section C.3) were tested for using discrimination accuracy, reaction time, subbing, behavioural inattention, and unstructured symbol cancellation. Visuospatial memory (Section C.4) was evaluated using abstract figures retrieval scores and verbal memory (Section C.5) for listed word recognition scores. A pain scale during the MRI scanning was recorded (Section C.6) and a stroke sickness questionnaire administrating, investigating physical and psychosocial daily sickness (Section C.7).

Disconnectome. The probability of white matter disconnections caused by the stroke event was quantified accounting for controls' connectivity where the lesion occurred. Stroke lesions were manually delineated in T1-weighted MRI scans and subsequently normalized to the MNI152 space (2 mm resolution). The BCBtoolkit "normalization tool" was used with the enantiomorphic normalization option (http://toolkit.bcblab.com) [44]. From the Human Connectome Project (HCP), 7T MRI diffusion-weighted scans were processed for N=163 healthy participants, 45% males. For the healthy participants, whole-brain tractography was reconstructed using the same procedure reported in [12]. Then, disconnectome profiles were processed with the BCBtoolkit [45]. HCP tractography was filtered considering only streamlines passing through each stroke lesion. The filter tractography was binarized and averaged

across the HCP participants. As a result, for each stroke patient a map of probability, ranging from 0 to 1, was obtained to quantify lesion disconnections.

Spatial embedding. Dimensionality reduction of patients' disconnectome was obtained using the UMAP method [31]. A non-linear embedding method that distributes data variability along major axes. Dataset 1 3-dimensional disconnectome maps were vectorised and imported as features of the embedding methods. As UMAP parameters, an approximation of 15 neighbours and a minimum 0.1 Euclidean distance was set to obtain a two-dimensional embedding of dataset 1. A space locally connected as Riemannian manifold that we addressed in the paper as disconnectome morphospace. The UMAP embedding transformation was stored as a Python object, using the Pickle library, to apply the same low-dimensional transformation further when new patients are imported into the model. Subsequently, to have positive coordinates with a zero origin, the maximum negative dataset 1 UMAP values across dimensions were added to shift the coordinate scales (Umap 1 and Umap 2).

Relationship to neuropsychological scores. Statistical correlations between patient localisation in the disconnectome morphospace and neuropsychological scores were conducted. Before the multiple regression formula, UMAP coordinates were converted into a 2D nifti image (260x260 matrix, 0.05 mm pixel size), and a Gaussian kernel spatial smoothing of 1 mm was applied (using FSL libraries https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/). This step was conducted to consider the uncertainty of UMAP coordinates and obtaining a spatial distribution of patient localisation in the disconnectome morphospace. Pixel-wise Pearson correlations between the patient probability of localisation and neuropsychological scores were conducted with iterative loops in Python (python numpy.corrcoef). Medium effect size correlation results only were considered informative (). Subsequently, since multiple clusters of voxels survived the threshold, a principal component analysis (PCA) has been run to compress the patient coordinate distribution variability. Three main principal components have been considered (Python sklearn.decomposition.PCA). Subsequently, patients' principal components have been entered, as dependent variables, in the multiple regression model (Python sklearn.linear_model.LinearRegression) to predict neuropsychological scores:

$$score_{predicted\ at\ 1\ year} = c + \sum_{i=1}^{3} w_i\ PCA_{patient\ score} \quad (1.1)$$

Where c is the intercept of the linear regression model, w_i are the model weights, and $PCA_{patient\ score}$ the model variables obtained as the inner product between the patient distribution of localisation and PCA components.

The multiple regression formula was trained with dataset 2 disconnectomes and validated using the independent dataset 3 cohort.

Accuracy of prediction was assessed by normalising the prediction error ($score_{measured} - score_{predicted}$) to the maximum score obtained in the neuropsychological evaluation:

$$score_{accuracy} = 1 - \frac{|score_{measured} - score_{predicted}|}{max(score_{measured})} \quad (1.2)$$

White matter atlas of neuropsychological components. In order to create a white matter atlas of the evaluated neuropsychological assessments, white matter disconnectomes (dataset 1) were correlated with patients' PCA scores, evaluated by running the prediction model on the dataset 1. The former disconnectome data were used in defining the UMAP space, whereas the latter model weights as variables of the multiple regression model to predict long term neuropsychological symptoms. Using *randomise* (FSL libraries) a generalized voxel-based linear regression model was run, with disconnectome maps as independent variables and PCA scores as dependent variables. To address the result of replicability this procedure was repeated twice, splitting the dataset 1 into two halves of n=666 subjects each.

The *randomise* T-maps obtained were used to calculate the correspondent effect size maps (f^2, python code reported in http://www.bcblab.com/BCB/Coding/Coding.html). For each neuropsychological score three principal component scores were evaluated and the maximum effect size across the components was considered. Subsequently, the highest effect size across neuropsychological assessments was reported in the white matter atlas summary map (FSL libraries *find_the_biggest* function). The replicability of the neuropsychology white matter atlas was quantified by means of Pearson correlations between the two summary maps.

DSD web application development. The DSD web application was built using the Django framework (https://www.djangoproject.com). This web framework allows database manipulation and is Python-based. The DSD frontend was created with standard Javascript and css templates; whereas the backend is hosted in a DigitalOcean webserver (https://www.digitalocean.com). Gunicorn and Ngnix are used for the web application live production.

Visualization.

A visualisation of the results was performed using Trackvis (http://trackvis.org), FSLeyes for imaging data, and Python matplotlib and seaborn libraries for scatter plots and matrices.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. All the neuropsychological score map used for defining the white matter atlas of neuropsychological components are freely available at https://neurovault.org/collections/11260/. The raw dataset imported in the BCBtoolkit software to calculate individual patient disconnectomes is available at https://www.humanconnectome.org (7 T diffusion data). In addition, processed data are available on request to the corresponding author.

Code availability. The code used in the analyses is available as part of the BCBtoolkit package http://toolkit.bcblab.com and the DSD web application http://disconnectomestudio.bcblab.com. Any additional information is available on request to L.T.

References

1.	Karnath, H.O., Ferber, S. & Himmelbach, M. Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature* **411**, 950-953 (2001).

2.	Verdon, V., Schwartz, S., Lovblad, K.O., Hauert, C.A. & Vuilleumier, P. Neuroanatomy of hemispatial neglect and its functional components: a study using voxel-based lesion-symptom mapping. *Brain* **133**, 880-894 (2010).

3.	Doricchi, F. & Tomaiuolo, F. The anatomy of neglect without hemianopia: a key role for parietal-frontal disconnection? *Neuroreport* **14**, 2239-2243 (2003).

4.	Bates, E. Language and the infant brain. *J Commun Disord* **32**, 195-205 (1999).

5.	Caplan, D., Hildebrandt, N. & Makris, N. Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain* **119 (Pt 3)**, 933-949 (1996).

6.	Graff-Radford, J.*, et al.* The neuroanatomy of pure apraxia of speech in stroke. *Brain Lang* **129**, 43-46 (2014).

7.	Goldenberg, G. & Spatt, J. The neural basis of tool use. *Brain* **132**, 1645-1655 (2009).

8.	Manuel, A.L.*, et al.* Inter- and intrahemispheric dissociations in ideomotor apraxia: a large-scale lesion-symptom mapping study in subacute brain-damaged patients. *Cereb Cortex* **23**, 2781-2789 (2013).

9.	Berti, A.*, et al.* Shared cortical anatomy for motor awareness and motor control. *Science* **309**, 488-491 (2005).

10.	Karnath, H.O., Baier, B. & Nagele, T. Awareness of the functioning of one's own limbs mediated by the insular cortex? *J Neurosci* **25**, 7134-7138 (2005).

11.	Forkel, S.J.*, et al.* Anatomical evidence of an indirect pathway for word repetition. *Neurology* **94**, e594-e606 (2020).

12.	Thiebaut de Schotten, M., Foulon, C. & Nachev, P. Brain disconnections link structural connectivity with function and behaviour. *Nat Commun* **11**, 5094 (2020).

13.	Mort, D.J.*, et al.* The anatomy of visual neglect. *Brain* **126**, 1986-1997 (2003).

14.	Corbetta, M.*, et al.* Common behavioral clusters and subcortical anatomy in stroke. *Neuron* **85**, 927-941 (2015).

15.	Von Monakow, C. *Die Lokalisation im Grosshirn und der Funktion durch kortikale Herde*, (Verlag von J.F. Bergmann, Wiesbaden, 1914).

16.   Feeney, D.M. & Baron, J.C. Diaschisis. *Stroke* **17**, 817-830 (1986).

17.   Carrera, E. & Tononi, G. Diaschisis: past, present, future. *Brain* **137**, 2408-2422 (2014).

18.   Fornito, A., Zalesky, A. & Breakspear, M. The connectomics of brain disorders. *Nat Rev Neurosci* **16**, 159-172 (2015).

19.   Fox, M.D. Mapping Symptoms to Brain Networks with the Human Connectome. *N Engl J Med* **379**, 2237-2245 (2018).

20.   Galton, F. Classification of Portraits. *Nature* **76**, 617-618 (1907).

21.   Mitteroecker, P. & Huttegger, S.M. The Concept of Morphospaces in Evolutionary and Developmental Biology: Mathematics and Metaphors. *Biological Theory* **4**, 54-67 (2009).

22.   Bonkhoff, A.K.*, et al.* Outcome after acute ischemic stroke is linked to sex-specific lesion patterns. *Nature Communications* **12**, 3289 (2021).

23.   Varoquaux, G. & Thirion, B. How machine learning is shaping cognitive neuroimaging. *Gigascience* **3**, 28 (2014).

24.   Ruffle, J.K.*, et al.* The autonomic brain: Multi-dimensional generative hierarchical modelling of the autonomic connectome. *Cortex* **143**, 164-179 (2021).

25.   Ardila, D.*, et al.* End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* **25**, 954-961 (2019).

26.   Oren, O., Gersh, B.J. & Bhatt, D.L. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *Lancet Digit Health* **2**, e486-e488 (2020).

27.   Wafa, H.A.*, et al.* Burden of Stroke in Europe: Thirty-Year Projections of Incidence, Prevalence, Deaths, and Disability-Adjusted Life Years. *Stroke* **51**, 2418-2427 (2020).

28.   Xu, T., Rolf Jager, H., Husain, M., Rees, G. & Nachev, P. High-dimensional therapeutic inference in the focally damaged human brain. *Brain* **141**, 48-54 (2018).

29.   Thiebaut de Schotten, M.*, et al.* From Phineas Gage and Monsieur Leborgne to H.M.: Revisiting Disconnection Syndromes. *Cerebral Cortex* **25**, 4812-4827 (2015).

30.   Vu, A.T.*, et al.* High resolution whole brain diffusion imaging at 7T for the Human Connectome Project. *Neuroimage* **122**, 318-331 (2015).

31.   McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **1802.03426v2** (2018).

32.   Rojkova, K., *et al.* Atlasing the frontal lobe connections and their variability due to age and education: a spherical deconvolution tractography study. *Brain Struct Funct* **221**, 1751-1766 (2016).

33.   Catani, M. & Thiebaut de Schotten, M. *Atlas of Human Brain Connections*, (Oxford University Press, Oxford, 2012).

34.   Cao, J., *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).

35.   Packer, J.S., *et al.* A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution. *Science* **365**(2019).

36.   Sirsat, M.S., Ferme, E. & Camara, J. Machine Learning for Brain Stroke: A Review. *J Stroke Cerebrovasc Dis* **29**, 105162 (2020).

37.   Calamia, M., Markon, K. & Tranel, D. The robust reliability of neuropsychological measures: meta-analyses of test-retest correlations. *Clin Neuropsychol* **27**, 1077-1105 (2013).

38.   Forkel, S.J., *et al.* Anatomical predictors of aphasia recovery: a tractography study of bilateral perisylvian language networks. *Brain* **137**, 2027-2039 (2014).

39.   Umarova, R.M., *et al.* Predictors and signatures of recovery from neglect in acute stroke. *Ann Neurol* **79**, 673-686 (2016).

40.   Forkel, S.J., Friedrich, P., Thiebaut de Schotten, M. & Howells, H. White matter variability, cognition, and disorders: a systematic review. *Brain Struct Funct* (2021).

41.   Hillis, A.E., *et al.* Hypoperfusion of Wernicke's area predicts severity of semantic deficit in acute stroke. *Ann Neurol* **50**, 561-566 (2001).

42.   Jha, A., *et al.* Metabolic lesion-deficit mapping of human cognition. *Brain* **143**, 877-890 (2020).

43.   Forkel, S.J. & Catani, M. Lesion mapping in acute stroke aphasia and its implications for recovery. *Neuropsychologia* **115**, 88-100 (2018).

44.   Nachev, P., Coulthard, E., Jager, H.R., Kennard, C. & Husain, M. Enantiomorphic normalization of focally lesioned brains. *Neuroimage* **39**, 1215-1226 (2008).

45.   Foulon, C., *et al.* Advanced lesion symptom mapping analyses and implementation as BCBtoolkit. *Gigascience* **7**, 1-17 (2018).
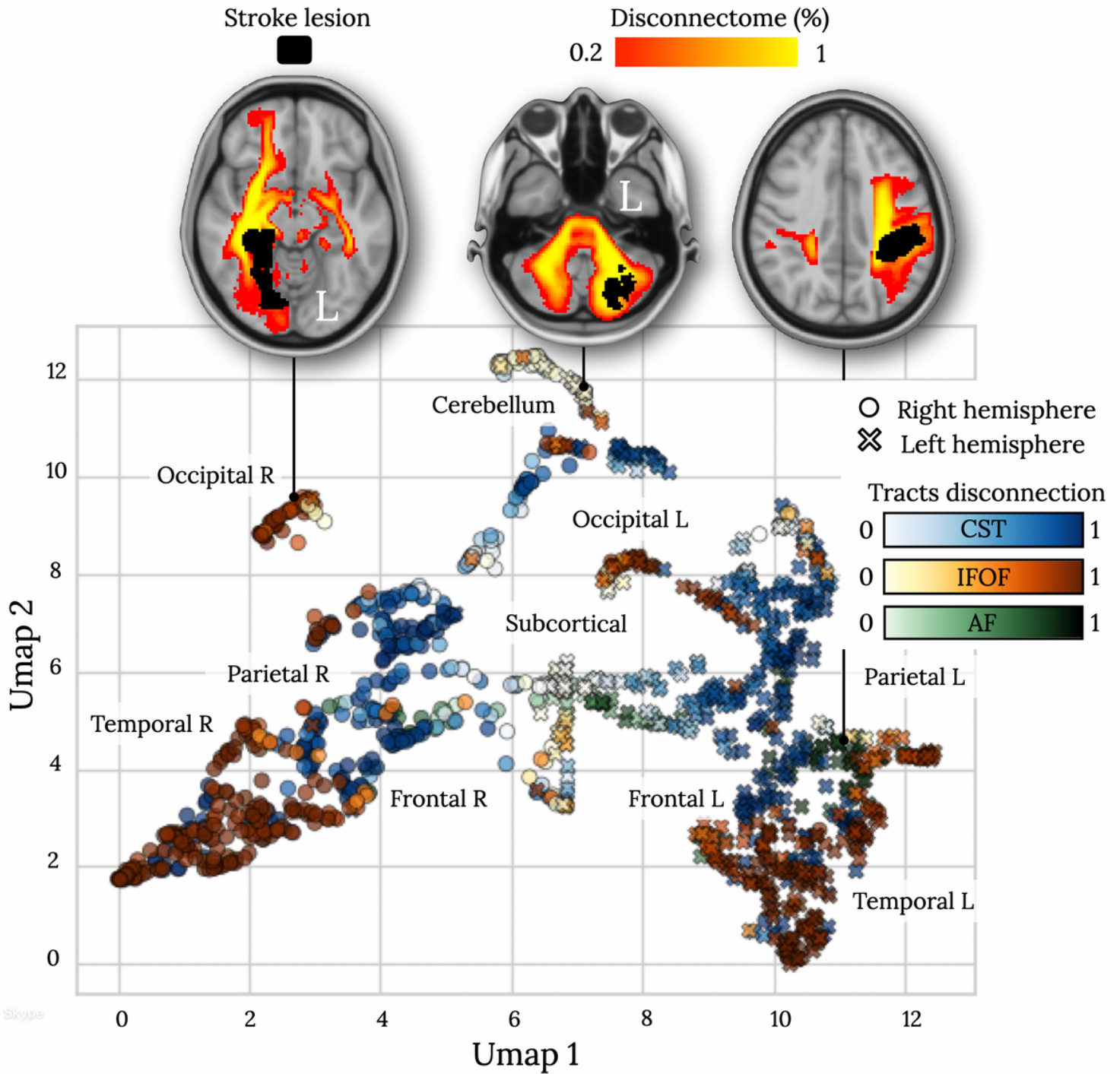
Figures

Figure 1

Disconnectome morphospace. Embedding of N=1333 stroke disconnectomes using Uniform manifold approximation and projection (Umap). (Top) three patients' disconnectomes, where a red-yellow colour map indicates the probability of disconnection. (Bottom) the disconnectome morphospace, where patients are marked with a circle ("o") or with a cross ("x") when their disconnection involves dominantly the right (R) or the left (L) hemisphere, respectively. Colours in the morphospace indicate the probability of disconnection of three white matter tracts: the cortico-spinal tract (CST), the arcuate fasciculus (AF),

and the inferior-fronto occipital fasciculus (IFOF). Brain lobe annotations indicate the dominant location of each stroke lesion.
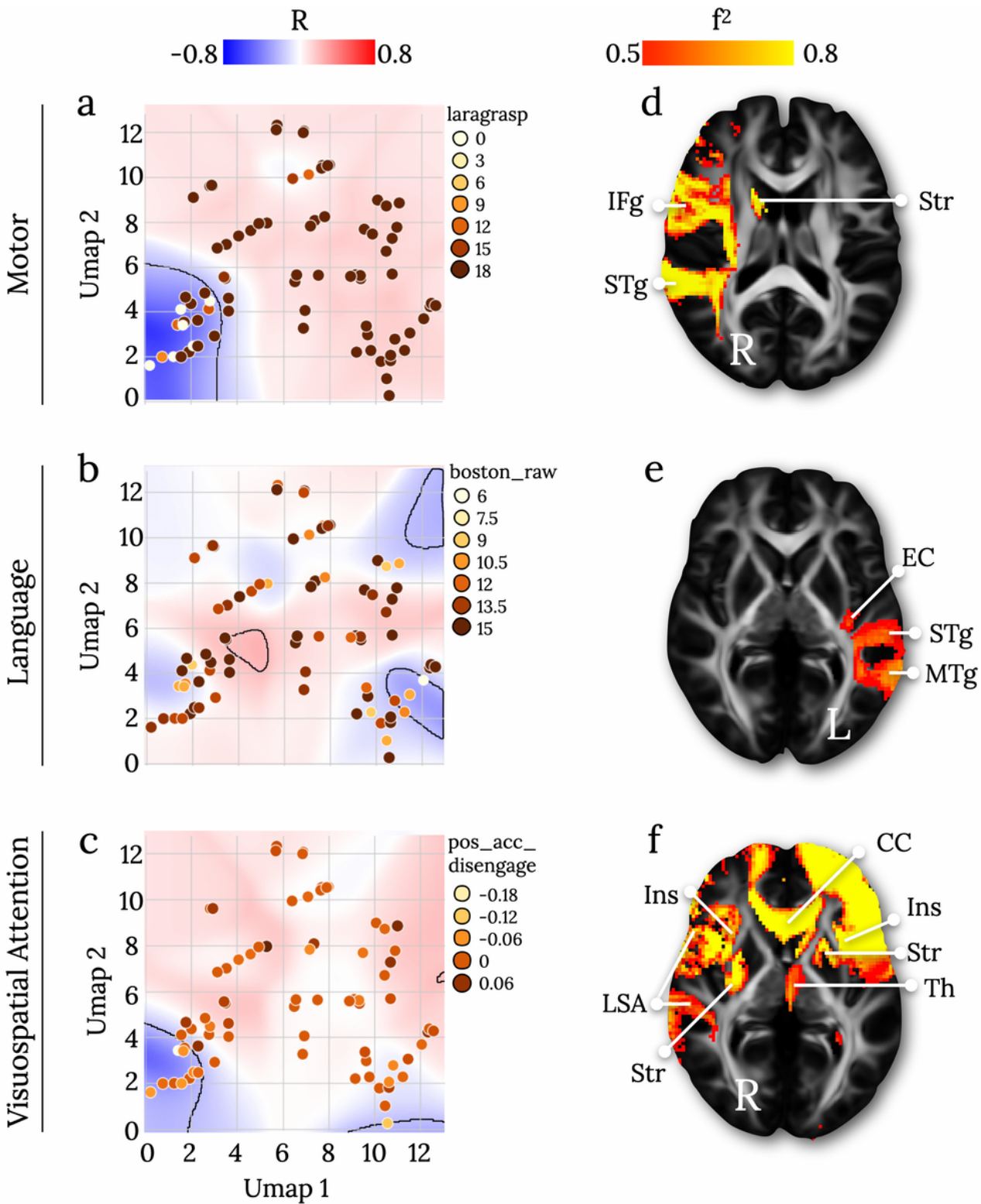

Figure 2

Composite morphospace. The composite morphospace corresponds to the disconnectome morphospace statistically combined with individual neuropsychological scores, (a-c) are three examples of different

neuropsychological score associations with morphospace territories presented together with (d-f) their prototypical disconnection profile. Blue-Red background colours in the morphospace correspond to Pearson correlation scores (R) with neuropsychological scores. Medium effect size territories (F) are delineated in black. All neuropsychological assessments and maps are reported in the Supplementary Materials (Section C). i: effect size; laragrasp: left grasping Action Research Arm test test; boston_raw: Boston naming test; pos_acc_disengage: accuracy in the Posner orienting task; CC: Corpus Callosum; EC: External/Extreme Capsule; IFg: Inferior Frontal gyrus; Ins: Insula; LSA: Long Segment of the Arcuate fasciculus; MTg: Middle Temporal gyrus; STg: Superior Temporal gyrus; Str: Striatum; Th: Thalamus.
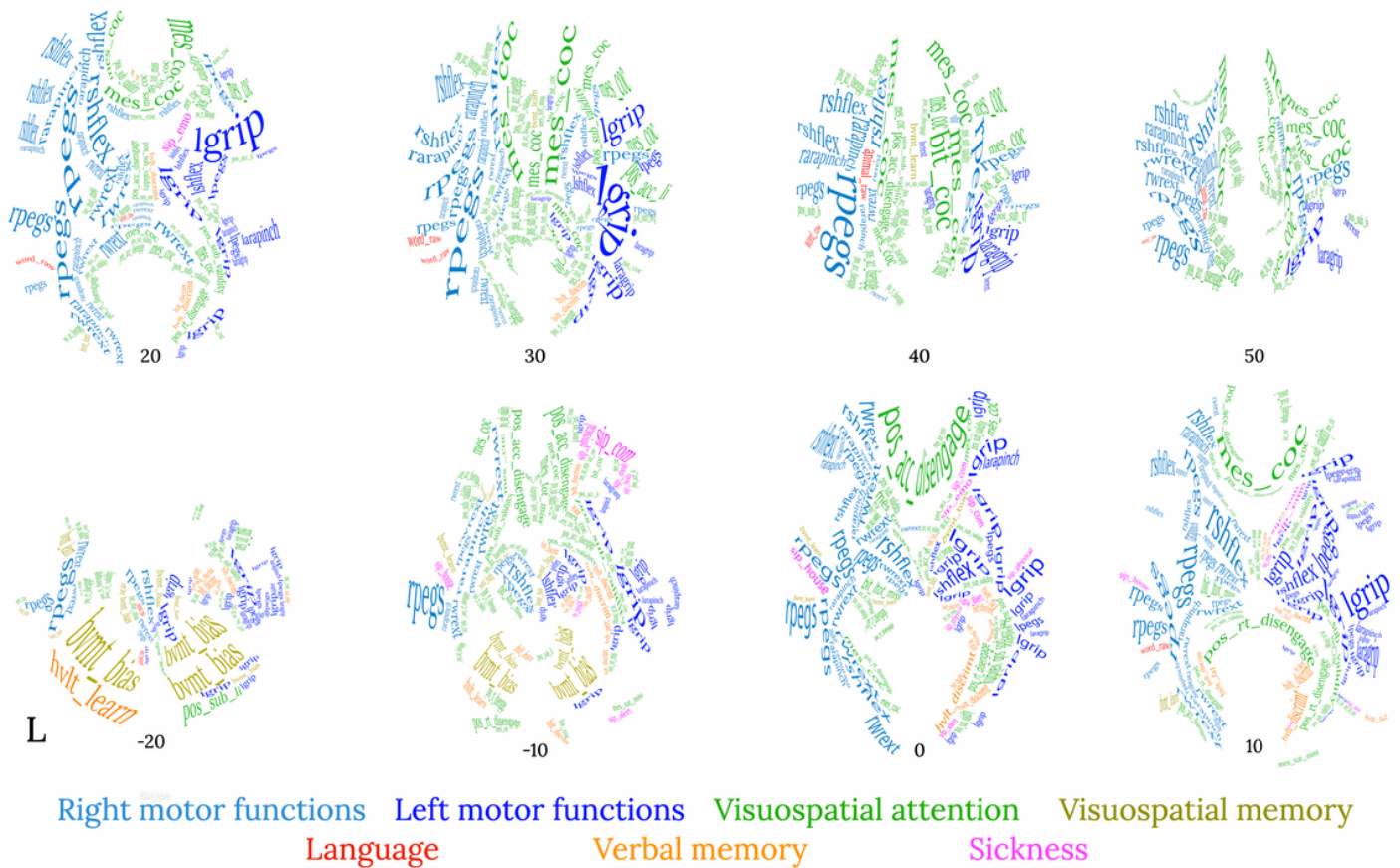

Figure 3

White matter neuropsychological atlas. Summary maps of neuropsychological scores associated with white matter disconnections with the highest statistical level (see Supplementary Materials Section D for high-resolution images).
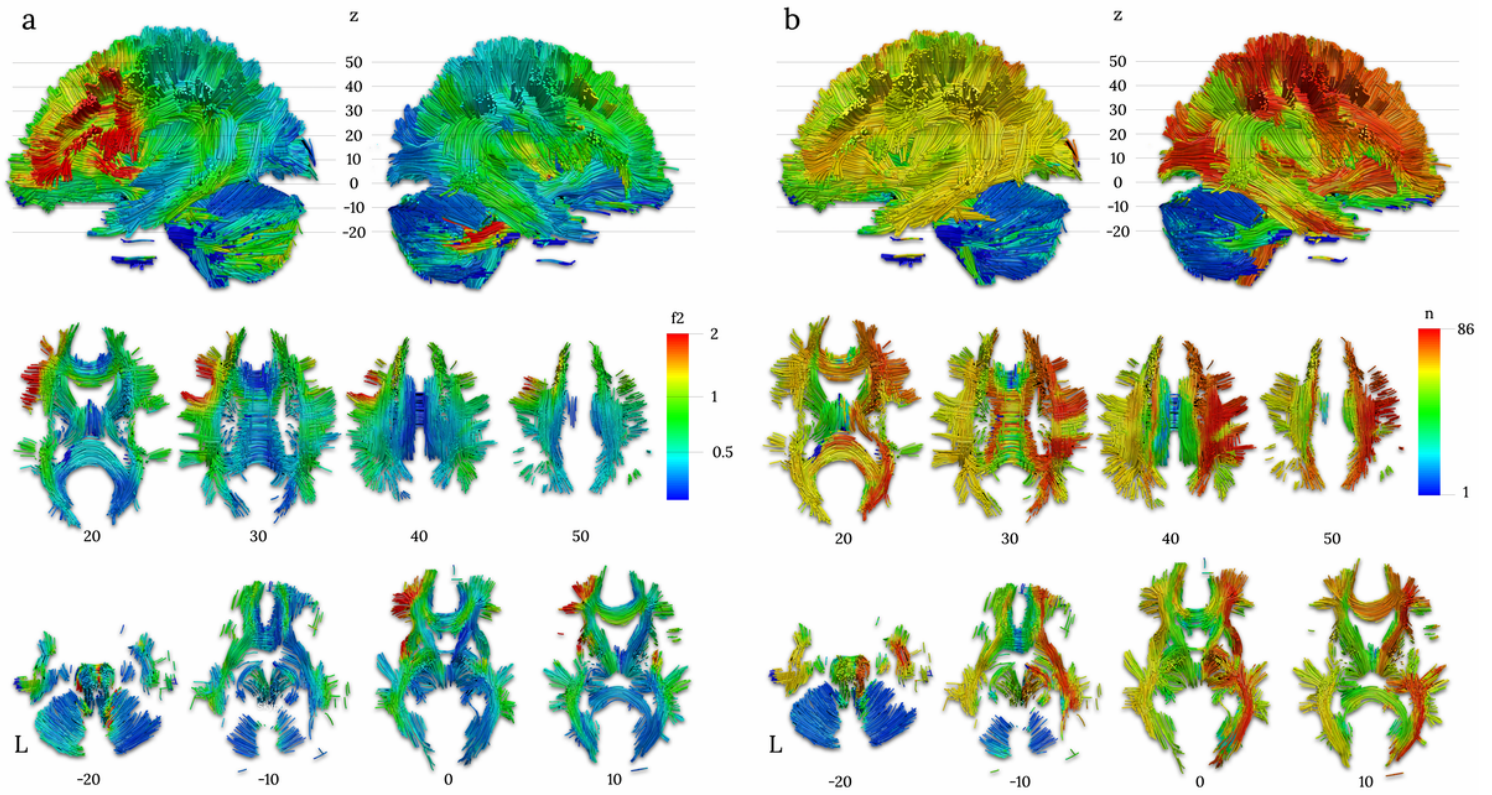
Figure 4

Neuropsychological score statistic onto white matter. (a) Effect size maps (f^2) for the most significant associations between disconnectome and neuropsychological performances. **(b)** Disconnectome versability maps indicating how many neuropsychological scores can be predicted for each voxel of white matter.
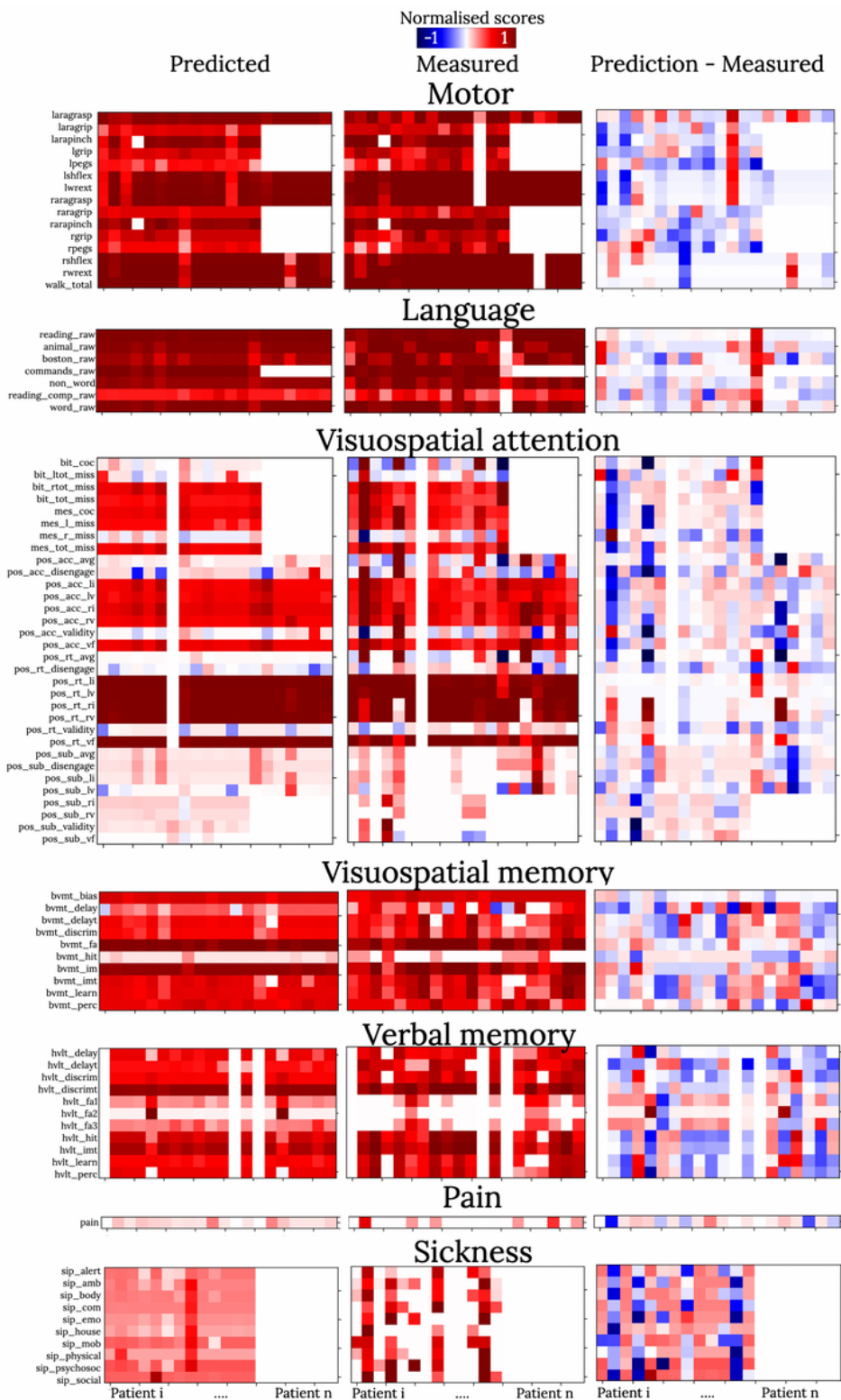
Figure 5

Accuracy of the neuropsychological scores. (left) Predicted neuropsychological scores, according to the composite disconnectome morphospace modelling. (center) Measured neuropsychological scores 1-year after the stroke onset. (right) Normalised error as the difference between predicted and measured scores. Rows correspond to single patients' neuropsychological profiles. Columns correspond to different

neuropsychological scores. See the Supplementary Materials Figure 4 for the same figure derived from the training set.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterialsDisconnectomeMorphospace.docx